

DRAFT 3 June 2011

NCRI-NCI Joint Conference | Biomedical Informatics Without Borders  
13-14 June 2011 | Wellcome Trust, London, UK

## **Standard and Vocabularies Theme**

The sessions for this theme will run from 11.00 a.m. on Monday to 2 p.m. on Tuesday. There will be four sessions, each running for approximately 90 minutes: the results of these will be presented as part of a final cross-theme plenary later on Tuesday afternoon.

In contrast to previous joint conferences, this year's meeting has been organized as an invitation-only workshop, with the participants working together in five separate groups – one group per theme – to produce a summary report of value to the whole meeting, as well as to the wider community. The other themes are “Sharing Data from Large-Scale Genomic Studies”, “Imaging”, “Use of Clinical Data for Research Purposes” and “Tooling and Technologies”. During the sessions, we will aim to identify opportunities toward the re-use, integration, representation, reasoning, development, education and training for biomedical standards and vocabularies in relevant scenarios and environments. We will look for realistic and pragmatic opportunities and the components of a short-term roadmap - reporting what is available now, and what can be achieved in the near future.

Some of the participants will be in a position to discuss needs and expectations, and present concrete examples of situations in which methods for semantic development could make a significant difference to the research that they conduct or support, or to the health services that they help to deliver. Others will be in a position to present extant as well as novel standards and ontologies, to report upon existing deployments, and to explore the potential for re-use, development, and integration.

We would be very grateful if you could have an appropriate presentation ready – 5 to 10 minutes, with or without slides – summarizing your interest in, or a particular contribution to, the discussion. A collaboration site, using IdeaScale, has been setup to solicit ideas and to begin a discourse prior, during and after the meeting as well (see <http://goo.gl/CoXUg>). Feel free to use this site or you may submit your presentation notes to us by email for us to collect and share during the meeting as well.

We will use these presentations to punctuate, inform, and drive the four sessions. We will not organize them into fixed slots, but rather allow as much time as seems appropriate for related discussion: if you have additional material that might support that discussion – further slides to show, or a demonstration to present – then this would be very welcome. We are working on methods to support and fully capture the discussions (scribing and recording) to improve retrieval of important concepts for development of an outcome paper following the event.

We have chosen an initial focus, a particular area of interest, for each of the sessions. Although there is time for

four sessions, the following five are provisional sub-themes which we will expect to adjust (scope, granularity) based on interest by those attending. We will also define these better by their component subtopics in a separate email and with your contributions during the coming week as well: 1) Biomedical Resource Representation and Profiling (including ontologies); (2) Principles for Adopting Standards and Vocabularies; (3) Interoperability Scenarios; (4) Data Representation, Serialization and Reasoning; 5) Cross Realm Harmonization. However, we anticipate changes in focus, and in the balance of discussion, as the meeting progresses – and will be ready to extend any particularly profitable discussion across session boundaries.

We would welcome any input that you might have before, as well as during, the meeting that would help us refine and extend this list of topics, and we look forward to seeing you in June!

Best wishes

Susanna and Stuart

Susanna-Assunta Sansone

[sa.sansone@gmail.com](mailto:sa.sansone@gmail.com)

@ Skype: susanna-a.sansone

Stuart Turner

[stuart@leafpath.org](mailto:stuart@leafpath.org)

+1.916.596.0255 (voice & text)

@ Skype: turner.stuart

## Appendix I

### Subthemes

1. Biomedical Resource Representation and Profiling
  1. Review of work by the Ontology Representation Working Group (NCRI, NCI, NCBO and OBO Foundry)
    1. Discuss related work (e.g. Biomedical Resource Ontology [Trish Whetzel, Nigam Shah] and the Information Artifact Ontology [Alan Ruttenberg])
    2. Catalogue of resources (e.g. BioSharing catalogue [Susanna Sansone])
    3. Review research collaboration/social networking features [Mike Conlon]
    4. Resource types
      1. Formal ontologies
      2. Knowledge representation systems that are not ontologies (controlled vocabularies, etc.)
      3. Minimal information checklists
      4. Exchange formats
      5. Documentation (normative, informative, companion artifacts to ontologies such as usage notes, supporting documents, etc.)
      6. Algorithms
      7. Discourse
  2. Principles for Adopting Standards and Vocabularies
    1. Principles for Adopting Standards and Vocabularies (this would roll in emerging policy such as the PCAST report, intellectual property issues, outreach, education, etc)
  3. Interoperability Scenarios (including methods to review, research and benchmark them; some examples such as the one Eric Prud'hommeaux brought up with ex vivo imaging and pathology imaging; or benchmarking realistic use cases and data sets in the different environments mentioned such as SemWeb, HL7, caBIG. This could also include the discussion mentioned by Dianne Reeves about concerns about capturing how people are using standards such as BRIDG, what does it mean to be compliant, governance work, bridging legacy data, systems largely dependent on HL7 v2 messaging, etc.)
  4. Data representation, detailed clinical models, serialization, and reasoning (would include issues discussed around complex data types, intermediate representations, ISO 11179, tying meaning to representation, novel approaches, some collaboration with Tooling and Technologies group, etc.)
  5. Cross-realm harmonization - presentation of work regarding standards development in different research, clinical and geographic domains with emphasis on overlapping work and between jurisdictional and organizational boundaries (e.g. Pistoia, W3C HCLS, HL7 international